Effective Population Size and Population Subdivision in Demographically Structured Populations

Valérie Laporte[1] and Brian Charlesworth

Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom

Manuscript received April 2, 2002
Accepted for publication June 18, 2002

ABSTRACT

A fast-timescale approximation is applied to the coalescent process in a single population, which is demographically structured by sex and/or age. This provides a general expression for the probability that a pair of alleles sampled from the population coalesce in the previous time interval. The effective population size is defined as the reciprocal of twice the product of generation time and the coalescence probability. Biologically explicit formulas for effective population size with discrete generations and separate sexes are derived for a variety of different modes of inheritance. The method is also applied to a nuclear gene in a population of partially self-fertilizing hermaphrodites. The effects of population subdivision on a demographically structured population are analyzed, using a matrix of net rates of movement of genes between different local populations. This involves weighting the migration probabilities of individuals of a given age/sex class by the contribution of this class to the leading left eigenvector of the matrix describing the movements of genes between age/sex classes. The effects of sex-specific migration and nonrandom distributions of offspring number on levels of genetic variability and among-population differentiation are described for different modes of inheritance in an island model. Data on DNA sequence variability in human and plant populations are discussed in the light of the results.

IN an ideal (Wright-Fisher) population of N breeding adult diploid individuals, the rate of genetic drift is equal to $1/(2N)$ (Wright 1931). Natural populations, however, depart from many of the assumptions of the ideal population, notably panmixia and a binomial distribution of the number of successful offspring per parent. The effective population size, N_e, can be used to describe the rate of genetic drift in such populations and replaces the ideal population size N in the resulting formulas (Wright 1931; Wang and Caballero 1999). For example, under the infinite-sites model of neutral variability, widely used in molecular population genetics, we can write $\theta = 4 N_e \mu$ for the scaled mutation rate, where μ is the neutral mutation rate per nucleotide site (Kimura 1971).

Population subdivision is a major factor that can influence N_e and has attracted much attention. Wright (1943) derived an expression for the effective size of the whole set of populations in an island model, where each local population has an equal probability of contributing to the pool of migrants. He showed that, in this case, subdivision increases N_e for the whole population over that under panmixia. Since this pioneering work, additional factors (*e.g.*, non-Poisson offspring number distributions, different modes of inheritance, sex-specific migration rates) have been included in the island model, and other migration models have been studied, such as stepping-stone models and metapopulations with local extinction and recolonization events (Whitlock and Barton 1997; Wang and Caballero 1999).

The total effective size of a subdivided population can be expressed as a function of the mean within-deme effective population size and of F_{ST} (Wright 1951), which describes the amount of between-deme differentiation relative to the variability in the set of populations as whole. In contrast to the results for the island model, population structure may reduce the total effective population size below that for panmixia when different demes make unequal contributions to the migrant pool (Whitlock and Barton 1997). A difficulty is that effective population size is defined somewhat loosely as the size of an ideal population with the same properties as the nonideal study population with respect to the effects of genetic drift (Wright 1931). Different effective population sizes can, however, be defined according to the variable of interest (Whitlock and Barton 1997) and do not necessarily give exactly equivalent results.

In particular, the inbreeding effective size of a subdivided population determines the asymptotic rate of the increase of the probability of identity of a pair of alleles sampled randomly from the population at large. The variance effective size determines the variance of the change in allele frequency across a generation for genes drawn from the population as whole. With constant population sizes across generations, these two parameters converge to the same value (Whitlock and Barton

[1]*Corresponding author:* Institute of Cell, Animal and Population Biology, University of Edinburgh, W. Mains Rd., Edinburgh EH9 3JT, UK. E-mail: valerie.laporte@ed.ac.uk

1997). Similarly, the mutation effective size is defined as the population size that gives the average pairwise genetic diversity between alleles sampled randomly from the population as a whole, when substituted into the above equation for θ (Whitlock and Barton 1997). The migration effective size for a subdivided population determines a weighted mean pairwise nucleotide site diversity for alleles sampled from the same population (Nagylaki 1998a).

These measures of effective population size were originally derived from recursion equations for identity probabilities of pairs of alleles or variances in allele frequencies (reviewed by Wang and Caballero 1999). More recently, use has been made of the coalescent process, which is very useful for studying alleles sampled in a prescribed way from the set of local populations (Hudson 1990; Slatkin 1991; Wilkinson-Herbots 1998; Wakeley 1999, 2001). This approach has been applied mostly to models of autosomal loci, in which sex-specific migration, and departures from the assumptions of the ideal population for individual demes, have been ignored, although Rousset (1999) introduced a coalescence approach to the study of between-deme differentiation at autosomal loci with sex-specific migration rates.

There is, however, increasing interest in comparing the patterns of genetic variation and population differentiation for genes with different modes of inheritance (Ennos 1994; Hurles and Jobling 2001). This is particularly important in the context of studies of sex chromosome evolution, in which it is of interest to compare variability at homologues on the *X* and *Y* chromosomes (Bachtrog and Charlesworth 2000, 2002; Filatov *et al.* 2000, 2001). The expectation is that selective forces acting on a nonrecombining genome can greatly reduce the level of variability on an evolving *Y* or neo-*Y* chromosome (Charlesworth and Charlesworth 2000), but this effect may be confounded with the effects of demographic differences between males and females (Charlesworth 2001) or of sex-specific migration in a subdivided population (Filatov *et al.* 2001). These effects have primarily been investigated using the classical recursion equation approach, either for the effects of demography (Charlesworth 2001) or for modeling sex-specific migration under various modes of inheritance (Chesser *et al.* 1993; Chesser and Baker 1996; Wang 1997a,b, 1999).

Here we first present a general framework for studying the pairwise coalescent times and identity probabilities of genes with different modes of inheritance in a metapopulation consisting of demes connected by an arbitrary migration scheme. Individuals within each deme are divided into different age and sex classes, with different demographic properties and migration probabilities. By using the "fast-timescale" approximation of Nordborg (1997) for the flow of genes among different classes within demes, it is possible to reduce the complexities of this general model to a manageable size. This yields simple expressions for the effective sizes of single populations and the expected levels of genetic diversity between and within demes. We provide some explicit results for the case of an island model with equal deme sizes and review data on human and plant populations in their light.

A GENERAL MODEL OF MIGRATION AND DRIFT IN STRUCTURED POPULATIONS

We assume that individuals within a local population (deme) can be classified into different classes, *e.g.*, with respect to age and/or sex. Let f_{ijrs} be the probability of identity of a pair of genes sampled from an individual of class r in deme i and an individual of class s in deme j. We can write m_{ijru} for the probability that a gene sampled from an individual of class r in deme i originated from an individual of class u in deme j.

When the probability per time interval of mutation of an allele to a new state, μ, is independent of allelic state, class, and deme, f_{ijrs} is equivalent to the moment-generating function for the time to coalescence of a pair of genes of the specified type, with parameter -2μ (Hudson 1990; Wilkinson-Herbots 1998). (To allow for the possibility of age structure, the time interval does not necessarily correspond to one generation.) In particular, their mean time to coalescence, T_{ijrs}, is equal to the derivative of f_{ijrs} with respect to -2μ at $\mu = 0$. As usual, we assume that μ is sufficiently small that second-order terms can be neglected. We later consider the case of sex-specific mutation rates, using mean coalescence times derived in this way; this is legitimate, since coalescence times are independent of the mutation model.

Assuming that different individuals migrate independently of each other, and using the standard approach to modeling a geographically structured population (Nagylaki 1982, 1998a), we can write the equation for equilibrium,

$$f_{ijrs} = (1 - 2\mu)\left\{\sum_{k\neq l}\sum_{uv} m_{ikru}m_{jlsv}f_{kluv} + \sum_{k}\sum_{uv} m_{ikru}m_{jksv}\left[P^{ijk}_{rsuv} + (1 - P^{ijk}_{rsuv})f_{kkuv}\right]\right\}, \quad (1)$$

where P^{ijk}_{rsuv} is the probability of coalescence over one time interval of two genes sampled from individuals of class r in deme i and s in deme j, derived from individuals from deme k who belong to classes u and v, respectively (note that $P^{ijk}_{rsuv} = 0$ unless $u = v$).

For $i = j$ and $r = s$, and if migration involves individuals rather than gametes, we should strictly include a term on the right-hand side that covers the case when the two genes sampled come from the same migrant individual. The approximations leading to Equation 3 mean that this is unnecessary, since ignoring it introduces only a second-order term into the final result. (This is because the chance that a pair of genes from the same deme are sampled from the same individual

is of the order of the reciprocal of the deme size, and we ignore products of this and the migration probabilities.)

Writing $h_{ijrs} = 1 - f_{ijrs}$ for the probability of nonidentity of a pair of alleles, Equation 1 can conveniently be rewritten as

$$h_{ijrs} = 2\mu + (1 - 2\mu)\left\{\sum_{uv}[\sum_{kl} m_{ikru}m_{jlsv}h_{kluv} - \sum_{k} m_{ikru}m_{jksv}P^{ijk}_{rsuv}h_{kkuv}]\right\}. \quad (2)$$

To make further progress, we assume that migration rates between demes are of the same order as the probabilities that two genes sampled from the same deme coalesce over one time unit and that both are sufficiently small that second-order terms in these, as well as mutation rates, can be neglected. This assumption is commonly made in applications of coalescent theory (Hudson 1990; Wilkinson-Herbots 1998); if migration rates are much greater than these coalescent probabilities, little differentiation between demes is possible, so that this is not a particularly restrictive assumption.

To this order of approximation, the last term in braces on the right-hand side of (2) is nonzero only when $i = j = k$, corresponding to the case of two genes drawn from the same deme,

$$h_{ijrs} \approx 2\mu + (1 - 2\mu)\sum_{uv}\sum_{kl} m_{ikru}m_{jlsv}h_{kluv} - \delta_{ij}\sum_{u} P^{i}_{rsu}h_{ijuu}, \quad (3)$$

where δ_{ij} is the Kronecker delta ($\delta_{ij} = 1$ when $i = j$ and is otherwise 0); P^{i}_{rsu} is the probability that two genes sampled from deme i from individuals belonging to classes r and s coalesce in an individual of class u in the previous time interval, treating the deme as a closed population.

To obtain manageable results, we simplify further by removing the dependence of h_{ijrs} on r and s. We can average h_{ijrs} values over classes within demes by using a weight of α_{ir} for the contribution from class r in the ith deme. α_{ir} is chosen to be equal to the rth element of the left leading eigenvector of the matrix that describes the flow of genes among classes within deme i, scaled such that the elements of this eigenvector sum to one. This element represents the stationary-state probability that a randomly sampled gene from deme i originates from class r (Nordborg 1997). If the convergence to this stationary state is much faster than the rate of coalescence and migration, α_{ir} gives an accurate approximation to the probability of origin of the sampled gene (Nordborg 1997); this is the case, for example, for a discrete generation population, when the classes refer to different sexes. This also implies that the h_{ijrs} values can be treated as approximately independent of r and s; *i.e.*, they can be equated to a common term, h_{ij}. This is equivalent to the fast-migration limit of Nagylaki (1980) for a geographically structured population.

We can then conveniently define the net probability of origin from deme k of a gene sampled from deme i, m_{ik}, as

$$m_{ik} = \sum_{ru}\alpha_{ir}\alpha_{ku}m_{ijru}. \quad (4)$$

Substituting into Equation 3 and summing $\alpha_{ir}\alpha_{js}h_{ijrs}$ over r and s, we obtain

$$h_{ij} = \sum_{rs}\alpha_{ir}\alpha_{is}h_{ijrs} \approx 2\mu + (1 - 2\mu)\sum_{kl} m_{ik}m_{jl}h_{kl} - \delta_{ij}P_i h_{ii}, \quad (5)$$

where P_i is the probability of coalescence of a pair of genes sampled randomly from deme i, such that

$$P_i = \sum_{rsu}\alpha_{ir}\alpha_{is}P^{i}_{rsu}. \quad (6)$$

Using the argument of Charlesworth (2001), we can write

$$P_i = \frac{1}{2t_i N_{ie}}, \quad (7)$$

where t_i is the generation time of individuals from deme i and N_{ie} is the deme's effective population size. In the case of discrete generations, $t_i = 1$ for all demes.

By differentiating h_{ij} with respect to -2μ, we obtain the following expression for the mean coalescent time for a pair of alleles sampled from demes i and j (Hudson 1990):

$$T_{ij} \approx 1 + \sum_{kl} m_{ik}m_{jl}T_{kl} - \delta_{ij}P_i T_{ii}. \quad (8)$$

This provides a general set of linear relations that can be used to solve for the T_{ij}, to the assumed order of approximation, for any specific model. Some examples are described below. Higher moments of the distribution of pairwise coalescence times can similarly be obtained by successive differentiation of Equation 5 (Wilkinson-Herbots 1998).

It is also useful to apply the method of Nagylaki (1998a) for determining a weighted mean coalescent time for a pair of genes sampled from the same deme. The migration rates defined by the right-hand side of Equation 4 are identical in properties to the components of the standard genic migration matrix. Under the usual assumptions about this matrix (Nagylaki 1980, 1982, 1998a), it has a leading left eigenvector, ν, whose elements can be assumed to sum to one without loss of generality. Using the argument of Nagylaki (1998a), this suggests the definition of a weighted mean coalescence time, T_0, for a pair of alleles sampled from the same deme as

$$T_0 = \frac{\sum_i \nu_i^2 P_i T_{ii}}{\sum_i \nu_i^2 P_i} = \frac{1}{\sum_i \nu_i^2 P_i}, \quad (9)$$

where the coalescent time for each deme is weighted by the product of the reciprocal of the coalescence probability and the square of the corresponding component of ν. In the discrete generation case, T_0 corresponds to twice the "migration effective size" as defined by Nagylaki (1980, 1982, 1998a) and reduces to the standard expression for this quantity in the case of

Wright-Fisher sampling within demes of size N_i, for which $P_i = 1/(2N_i)$.

COALESCENT PROBABILITIES AND EFFECTIVE POPULATION SIZES

Separate sexes with discrete generations: The utility of the approach described above to determining the effective population size of a deme can be illustrated by its application to the discrete-generation case with separate sexes, previously treated by Caballero (1995). It enables us to provide biologically explicit expressions for N_e under different modes of inheritance, on the lines developed by Charlesworth (2001) for the age-structured case.

From the assumption that the products of coalescence probabilities and migration probabilities are negligible, used to derive (3) and its successors, it is legitimate to treat each deme as a closed population. We have to consider the origins of genes sampled from individuals of all permissible combinations of sexes and parents (which combination is permissible depends on the mode of genetic transmission). For a pair of maternally derived genes sampled from two different females, the probability that they both come from the same mother is approximately

$$Q^i_{\mathrm{fff}} = \frac{\sum_k d_{ik}(d_{ik} - 1)}{(N_{if}\bar{d}_i)^2} \tag{10}$$

(Caballero 1995), where N_{if} is the number of breeding females in deme i, d_{ik} is the number of daughters produced by the kth female in deme i, and $\bar{d}_i$ is the mean number of daughters per female in deme i.

This can be simplified by noting that the Poisson value for the variance in numbers of daughters per female, V^i_{ff}, is $\bar{d}_i$, so that in general the deviation of this variance from the Poisson value can be written as

$$\Delta V^i_{\mathrm{ff}} = V^i_{\mathrm{ff}} - \bar{d}_i \tag{11}$$

so that

$$Q^i_{\mathrm{fff}} = \frac{(\Delta V^i_{\mathrm{ff}}/\bar{d}_i^2) + 1}{N_{if}}. \tag{12}$$

Similar expressions can be obtained for other pairs of individuals and their parents (see Table 1).

Let β_{irsu} be the probability that a pair of genes sampled from individuals of sexes r and s in deme i both come from parents of sex u. Let γ_{irsu} be the probability that this pair of genes shared a common ancestral allele in the previous generation, given that they have a parent in common; *i.e.*, they coalesce. From Equation 7, the net probability that a randomly chosen pair of genes coalesce in the previous generation is

$$\frac{1}{2N_{ie}} = P_i = \sum_{rsu} \alpha_{ir}\alpha_{is}\beta_{irsu}\gamma_{irsu}Q^i_{rsu}, \tag{13}$$

TABLE 1

Origins of genes sampled from a given deme

Sexes of sampled individuals	Sex of parent	Probability of parent of given sex
Female, female	Female	$Q^i_{\mathrm{fff}} = \dfrac{(\Delta V^i_{\mathrm{ff}}/\bar{d}_i^2) + 1}{N_{if}}$
Male, male	Female	$Q^i_{\mathrm{mmf}} = \dfrac{(\Delta V^i_{\mathrm{fm}}/\bar{s}_i^2) + 1}{N_{if}}$
Male, female	Female	$Q^i_{\mathrm{fmf}} = \dfrac{(\Delta C^i_{\mathrm{ffm}}/\bar{d}_i\bar{s}_i) + 1}{N_{if}}$
Female, female	Male	$Q^i_{\mathrm{ffm}} = \dfrac{(\Delta V^i_{\mathrm{mf}}/\bar{d}_i^{*2}) + 1}{N_{im}}$
Male, male	Male	$Q^i_{\mathrm{mmm}} = \dfrac{(\Delta V^i_{\mathrm{mm}}/\bar{s}_i^{*2}) + 1}{N_{im}}$
Male, female	Male	$Q^i_{\mathrm{fmm}} = \dfrac{(\Delta C^i_{\mathrm{mfm}}/\bar{d}_i^*\bar{s}_i^*) + 1}{N_{im}}$

i indicates the identity of the deme. ΔV^i_{rs} is the excess over the Poisson expectation (equal to the mean number of progeny per capita of sex s) of the variance of the number of progeny of sex s of parents of sex r.

where Q^i_{rsu} is the probability that a pair of genes sampled from individuals of sexes r and s in deme i with a parent of sex u derives from the same parental individual.

Table 2 gives the values of these transmission probabilities for different modes of inheritance; these can be combined with the Q values from Table 1 and substituted into Equation 15, to obtain the final expressions for effective population sizes. Simplified expressions can be obtained when there is a fixed sex ratio among breeding individuals, so that there is a binomial distribution of the proportions of sons and daughters of a given individual, counted at maturity (see the appendix).

For convenience, we henceforth drop the superscripts and subscripts indicating deme identities. Using Tables 1 and 2, substituting into Equation 15, and noting that the Poisson expectation for variance in total offspring number is $1/(1 - c)$ for females and $1/c$ for males, where c is the proportion of males among breeding adults, we obtain the expression for effective population size with autosomal inheritance,

$$\frac{1}{N_{\mathrm{eA}}} \approx \frac{(1 + F)}{4}\left\{\frac{1}{N_{\mathrm{f}}} + \frac{1}{N_{\mathrm{m}}} + \frac{(1 - c)^2\Delta V_{\mathrm{f}}}{N_{\mathrm{f}}} + \frac{c^2\Delta V_{\mathrm{m}}}{N_{\mathrm{m}}}\right\}, \tag{14}$$

where ΔV_{f} and ΔV_{m} are the excesses over Poisson expectation of the total numbers of offspring per capita for female and male parents, respectively; F is the inbreeding coefficient associated with departure from random mating within demes.

For X-linked inheritance with male heterogamety, we have

$$\frac{1}{N_{eX}} \approx \frac{1}{9}\left\{\frac{4(1+F)}{N_f} + \frac{2}{N_m} + \frac{4(1+F)(1-c)^2\Delta V_f}{N_f} + \frac{2c^2\Delta V_m}{N_m}\right\}. \quad (15)$$

For *Y*-linked inheritance

$$\frac{1}{N_{eY}} \approx \frac{2(1 + c^2\Delta V_m)}{N_m}. \quad (16)$$

With female heterogamety, males and females are interchanged in these expressions.

For maternal inheritance

$$\frac{1}{N_{eC}} \approx \frac{2(1 + [1-c]^2\Delta V_f)}{N_f}. \quad (17)$$

Partially self-fertilizing hermaphrodites: It is also of interest to consider the case of a nuclear gene in a population of partially self-fertilizing hermaphrodites; this is particularly relevant for plants. (Maternally transmitted genes are clearly unaffected by the selfing rate, except as far as changes in the breeding system alter the distribution of numbers of successful offspring produced through seed.) With a Poisson distribution of offspring number for both seed and pollen, the effective size is reduced below that for random mating by a factor of $1/(1 + F)$, where F is the equilibrium inbreeding coefficient produced by the level of self-fertilization (Pollak 1987; Nordborg and Donnelly 1997). We examine the effect on this result of deviations from a Poisson offspring number distribution.

Let there be N breeding individuals in the deme under consideration; the kth individual is assumed to produce d_{kS} progeny through self-fertilized seed that survive to form part of the breeding population next generation, and d_{kO} such progeny through outcrossed seed. In addition, it contributes p_k progeny through pollen that fertilizes other individuals. The overall probability that a progeny individual is the product of self-fertilization is then

$$S = \frac{\sum_i d_{iS}}{\sum_i (d_{iS} + d_{iO})} = \frac{\bar{d}_S}{(\bar{d}_S + \bar{d}_O)}, \quad (18)$$

where $\bar{d}_S$ and $\bar{d}_O$ are the mean numbers of successful progeny per capita produced by selfed and outcrossed seed, respectively.

Details of the derivation of probabilities of gene origins are given in the appendix. In the case of constant population size, we have $\bar{d} = 1$, $\bar{d}_S = S$, and $\bar{d}_O = \bar{p} = (1 - S)$. The expression for the effective population size for a single deme, N_{eH}, is

$$\frac{1}{N_{eH}} \approx \frac{(1+F)}{N}\left\{1 + \frac{(\Delta V_f + \Delta V_m + 3\Delta V_{fS}) + 2[C_{fOfS} + C_{fm} + C_{fSm}]}{4}\right\} \quad (19a)$$

(definitions of the terms in the numerator are given in the appendix).

If the number of successful offspring through selfed seed, conditioned on the total number of successful offspring produced through seed, can be regarded as a binomial variate with parameter S (*i.e.*, there is the same expected selfing rate for each individual), this simplifies further to

$$\frac{1}{N_{eH}} \approx \frac{(1+F)}{N}\left\{1 + \frac{([1+S]^2\Delta V_f + \Delta V_m + 2[C_{fm} + C_{fSm}])}{4}\right\}. \quad (19b)$$

In the limit of complete self-fertilization, the contributions from outcrossing pollen vanish, and $F = 1$, so that

$$\frac{1}{N_{eH}} \approx \frac{2(1 + \Delta V_f)}{N}. \quad (20)$$

A model is presented in the appendix for analyzing the case of intermediate selfing rates. The overall conclusion is that non-Poisson variation in reproductive capacity may not greatly alter the standard result that N_e for a selfing population is reduced by a factor of $1/(1 + F)$, relative to the value for an outcrossing population (Pollak 1987), although a very high sensitivity of female fitness to allocation to reproduction could lead to a greater reduction in N_e than this.

THE ISLAND MODEL WITH SEX-SPECIFIC MIGRATION PARAMETERS

In this section, we apply the above principles to the simple case of an island model (Wright 1943) with sex-specific migration rates, for dioecious populations under a variety of modes of inheritance. We assume that the total population is divided into n demes, each with the same effective population size, N_e, given by the equations derived above. There are three dispersal parameters, describing the probability that a male gamete (m_m) in a zygote in a given deme has migrated from a different deme, the corresponding probability for a female gamete (m_f), and the probability that a zygote is itself derived from a different deme (m_z; Chesser and Baker 1996; Wang 1999). Under the assumptions used above, these migration events can be treated as mutually exclusive, since their probabilities are each low. Migration probabilities between specific demes, as considered above, are given by dividing the dispersal parameters by $(n - 1)$.

In plant species, male gametes can disperse through pollen but female gametes cannot disperse, so that only m_m and m_z can be nonzero. For animals, migration of male and female gametes occurs via dispersal of males and unmated females; dispersal of zygotes occurs when females migrate after mating but before laying eggs or giving birth. All three migration parameters can therefore be nonzero.

Using the argument leading to Equation 4, we can then define the net migration rates for different modes of inheritance (autosomal, *X*-linked, *Y*-linked, and cytoplasmic) as m_A, m_X, m_Y, *and* m_C, respectively. We have

TABLE 2

Transmission parameters for different modes of inheritance

Sexes of sampled individuals (r, s)	Sex of parent (u)	Mode of inheritance: Autosomal $\alpha_{ii}\alpha_{is}$	Autosomal β_{irsu}	Autosomal γ_{irsu}	*X*-linkage $\alpha_{ii}\alpha_{is}$	*X*-linkage β_{irsu}	*X*-linkage γ_{irsu}	*Y*-linkage $\alpha_{ii}\alpha_{is}$	*Y*-linkage β_{irsu}	*Y*-linkage γ_{irsu}
Female, female	Female	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{(1+F_i)}{2}$	$\frac{4}{9}$	$\frac{1}{4}$	$\frac{(1+F_i)}{2}$	—	—	—
Male, male	Female	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{(1+F_i)}{2}$	$\frac{1}{9}$	1	$\frac{(1+F_i)}{2}$	—	—	—
Male, female	Female	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{(1+F_i)}{2}$	$\frac{4}{9}$	$\frac{1}{2}$	$\frac{(1+F_i)}{2}$	—	—	—
Female, female	Male	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{(1+F_i)}{2}$	$\frac{4}{9}$	$\frac{1}{4}$	1	—	—	—
Male, male	Male	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{(1+F_i)}{2}$	—	—	—	1	1	1
Male, female	Male	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{(1+F_i)}{2}$	—	—	—	—	—	—

F_i is the inbreeding coefficient due to consanguineous matings within deme i. Maternally transmitted organelle genomes have similar properties to *Y*-linked genes, except that female and male parameters are interchanged.

$$m_A = m_z + \frac{1}{2}(m_m + m_f) \tag{21a}$$

$$m_X = m_z + \frac{1}{3}(m_m + 2m_f) \tag{21b}$$

$$m_Y = m_z + m_m \tag{21c}$$

$$m_C = m_z + \frac{1}{2}m_f. \tag{21d}$$

Coalescence times and expected nucleotide site diversities: Results for mean pairwise coalescence times in the island model can be obtained by substituting these expressions into standard formulas (Slatkin 1991; Wilkinson-Herbots 1998), which correspond to approximate solutions to Equation 8. To the assumed order of approximation, with mode of inheritance G (where G is A, X, Y, or C) the expected coalescence time, T_{0G}, for two genes sampled from the same deme is given by

$$T_{0G} = 2nN_{eG}, \tag{22a}$$

where N_{eG} is the appropriate effective population size for a single deme.

Using Equation 8, or the argument of Slatkin (1991), the approximate expected coalescence time for a pair of genes sampled from different demes is

$$T_{1G} = T_{0G} + \frac{(n-1)}{2m_G}, \tag{22b}$$

where m_G is the migration rate defined by the relevant choice of Equations 21 for mode of inheritance G.

The expected coalescence time for a pair of genes sampled randomly from the set of populations is

$$T_G = \frac{1}{n}T_{0G} + \frac{(n-1)}{n}T_{1G}. \tag{22c}$$

The expected pairwise nucleotide site diversity under the infinite-sites model is given by the product of coalescence time and the mutation rate μ_G for the given mode of inheritance (Hudson 1990). From Equations 22, the expectations of nucleotide diversities for alleles sampled within demes, π_{SG}, and from the population as a whole, π_{TG}, are given, respectively, by

$$\pi_{SG} = 4nN_{eG}\mu_G \tag{23a}$$

$$\pi_{TG} = \pi_{SG}\left(1 + \frac{(n-1)^2}{n^2}\frac{1}{4m_GN_{eG}}\right). \tag{23b}$$

Mutation rates may differ among genetic systems because of differences in mutation rates between males and females (Hurst and Ellegren 1998; McVean 2000) or because of chromosome-specific mutation rates (McVean and Hurst 1997). In the former case, the overall mutation rate is given by weighting the sex-specific or chromosome-specific mutation rates by the appropriate α values (Charlesworth 2001).

The absolute magnitude of between-population subdivision can be measured by the difference between the nucleotide diversity for the population as a whole and the mean for a pair of alleles sampled from the same deme (Charlesworth 1998). The expectation of this

is given by the second term in brackets on the right of (23b). It is often more convenient to use a relative measure, such as F_{ST}. Using the definition of HUDSON *et al.* (1992), the theoretical value of F_{ST} for a given mode of inheritance is given by

$$F_{ST,G} = \frac{(\pi_{TG} - \pi_{ST})}{\pi_{TG}}. \quad (24a)$$

With a large number of demes, this yields the familiar result of WRIGHT (1951):

$$F_{ST,G} = \frac{1}{4N_{eG}m_G + 1}. \quad (24b)$$

We now consider the application of these formulas to some specific examples. We focus initially on the effects of population subdivision, assuming a 1:1 sex ratio, Poisson variances in fertility for both sexes, and equal mutation rates for males, females, and different chromosomes. Discrete generations and a deme size of N breeding adults for all demes are also assumed. In this case, the ratios of effective population sizes and within-deme diversities for different modes of inheritance G and G' ($r_{GG'}$) are simply equal to the ratios of the respective numbers of gene copies in the population: $r_{XA} = \frac{3}{4}$, $r_{XY} = r_{XC} = 3$, $r_{AY} = r_{AC} = 4$, the same values as for panmixia, which are often quoted in the literature on molecular population genetics.

The effects of sex-specific migration and population subdivision in dioecious plants: In Figure 1, we plot the values of F_{ST} for nuclear and cytoplasmic genes and the ratios of the total diversities for different modes of inheritance, as functions of the total number of migrants per generation (give by the sum of pollen and seed migration rates times the deme size). Male heterogamety is assumed.

As expected, reduced migration increases F_{ST} for all modes of inheritance. However, the rate of increase of F_{ST} as the amount of migration decreases is different for different modes of inheritance and also depends on the mode of migration. For instance, with equal pollen and seed migration rates, $F_{ST,Y}$ and $F_{ST,C}$ increase faster than $F_{ST,A}$ and $F_{ST,X}$ (Figure 1A). These differences result from differences in the numbers of effective migrants, due to two factors: (i) different effective population sizes and (ii) different gene migration rates. The effective number of X chromosomal migrants is lower than the effective number of autosomal migrants, due to a lower effective population size of the X chromosomes ($3N/4$) as well as its lower migration rate ($m_z + \frac{1}{3}m_m$ *vs.* $m_z + \frac{1}{2}m_m$). The effective number of Y chromosomal migrants is lower than both the effective number of chromosomal X or autosomal migrants, due to the lower effective population size of the Y chromosome ($N/4$), which is not compensated for by its higher migration rate ($m_z + m_m$) relative to the X chromosome. A similar argument applies to cytoplasmic genes.

Population subdivision also modifies the ratios of the total diversities for different modes of inheritance: the ratios $\pi_{TX}/\pi_{TY}(R_{XY})$ and $\pi_{TA}/\pi_{TY}(R_{AY})$ decrease sharply with reduced gene flow and $\pi_{TX}/\pi_{TA}(R_{XA})$ increases slightly (Figure 1D). From Equations 23, it is easily seen that, in the limit when the migration rates tend toward zero, the R values are equal to the reciprocals of the ratios of the respective migration rates given by Equations 21, since the right-hand side is dominated by the term involving the reciprocal of the product of effective size and migration rate. In this case, the limiting total diversities for autosomal and X-linked genes are only 1.33- and 1.5-fold higher than for Y-linked loci, respectively, with the diversity for X-linked loci being 1.12 that for autosomal loci. Similarly, the limiting ratio of autosomal to cytoplasmic total diversities becomes 0.67.

Differences between the pollen and seed migration rates also influence the rate of increase in F_{ST}. For instance, if migration occurs primarily through pollen (*e.g.*, with $m_m = 100m_z$, Figure 1B), $F_{ST,A}$ and $F_{ST,X}$ increase faster with decreasing migration than when migration occurs equally through pollen and seeds (Figure 1A), or primarily through seeds ($m_m = 0.01m_z$, Figure 1C), whereas $F_{ST,C}$ increases more slowly, but is higher throughout the range of Nm displayed. This is because X-linked and autosomal genes are haploid in pollen grains but diploid in seeds, and cytoplasmic genes move only through seed. In contrast, the Y chromosome, which is haploid in both dispersal units, is not affected by the relative pollen and seed migration rates. In consequence, the curves relating the diversity ratios to total migration rate are also modified. In particular, when migration occurs primarily through pollen, the increased migration rate of the Y chromosome compared to the X chromosome (m_m *vs.* $m_z/3$) compensates nearly totally for its reduced effective population size, so that R_{XY} is independent of population subdivision (Figure 1E). In contrast, R_{XA} is now slightly more affected by population subdivision, because the difference between the X and autosomal migration rates increases with the pollen migration rate. Hence, in the limit when the migration rates tend toward zero, R_{XY} barely declines below 3, but R_{AY} declines to 2, and R_{XA} increases to 1.5.

The effects of sex-specific migration and population subdivision in animals: We consider a model of animal migration with up to a sixfold difference between male and female migration rates and no zygotic migration (see above), assuming male heterogamety (for female heterogamety, male and female parameters are interchanged). As compared with the plant case, we can see that the case of predominantly male migration (Figure 2, B and E) is similar to the case of predominantly pollen migration in the plant model. The case of equal male and female migration rates is similar to the case of equal pollen and seed migration rates, except that limited migration has a much stronger effect on F_{STY} relative to F_{STA} and F_{STX} (Figure 2A). The effect is even stronger with

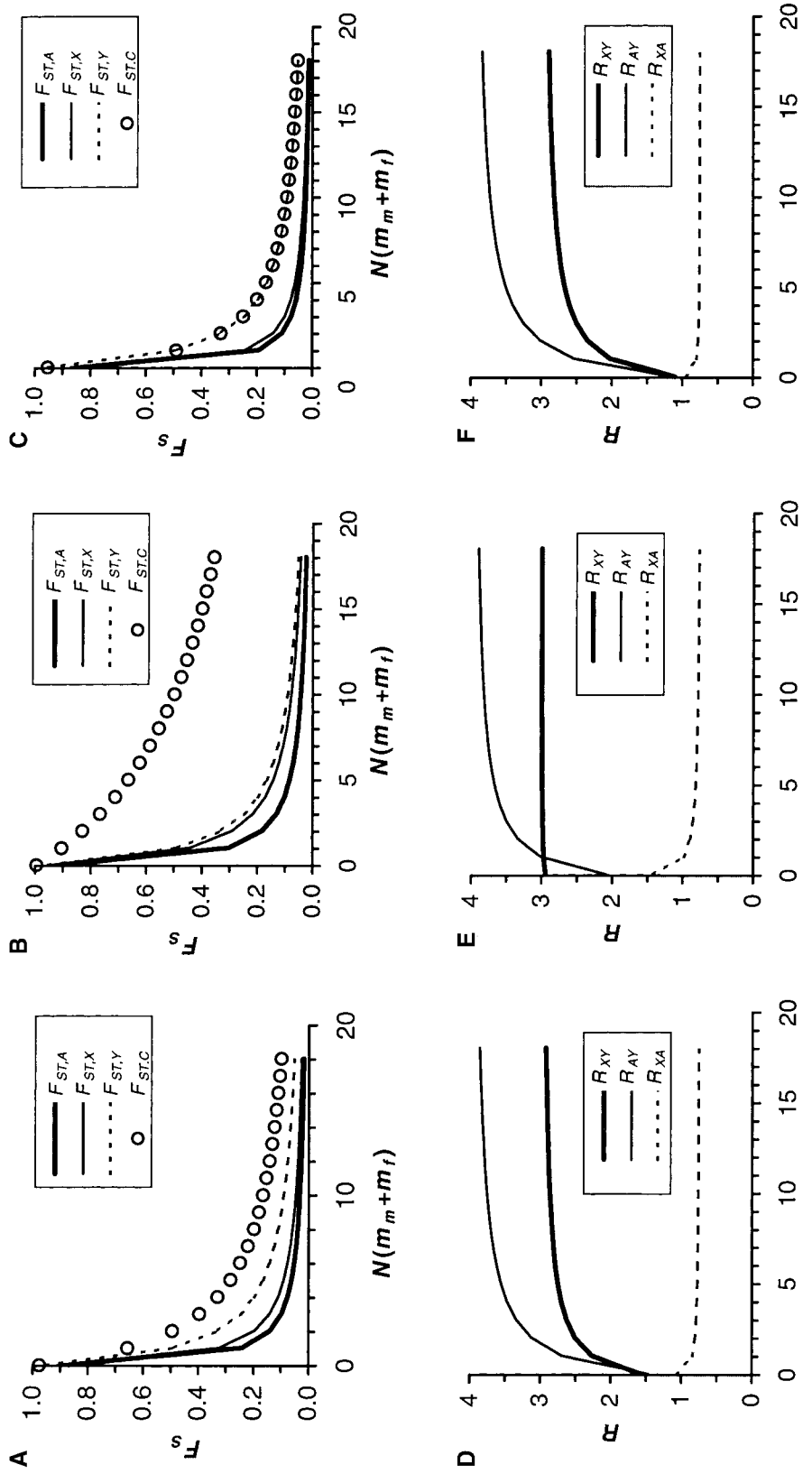

FIGURE 1.—The F_{ST} parameters (A–C) and the ratios (R) of total diversities between genes with different modes of inheritance (D–F) in a subdivided plant population ($m_f = 0$), as functions of the number of migrants per deme, $N(m_m + m_z)$. (A and D) Equal pollen and seed migration. (B and E) Predominantly pollen migration ($m_m = 100m_z$). (C and F) Predominantly seed migration ($m_z = 100m_m$). We assume $N_{eA} = N$, $N_{eX} = \frac{3}{4}N$, $N_{eY} = N_{ec} = \frac{1}{4}N$.

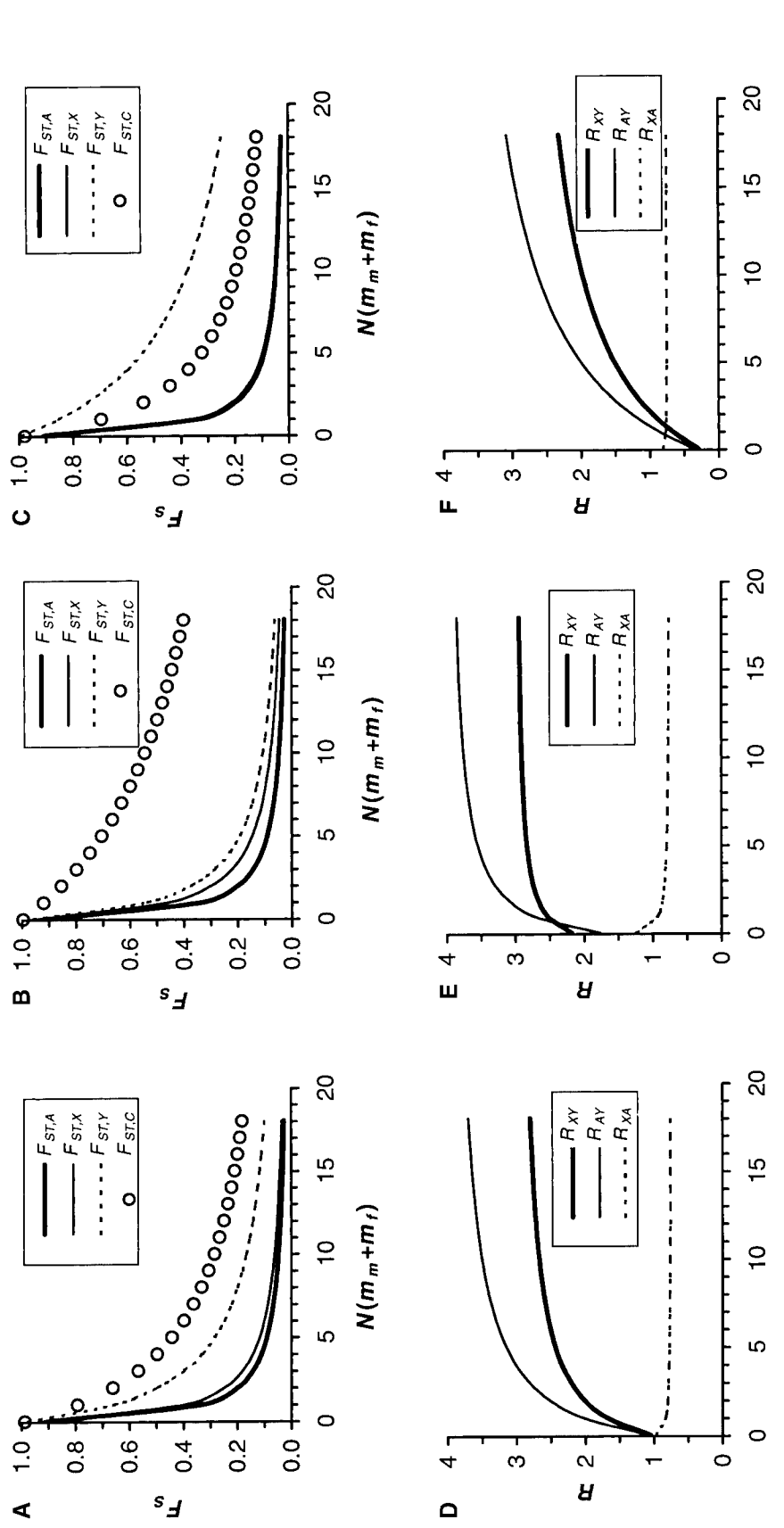

FIGURE 2.—The F_{ST} parameters (A–C) and the R values (D–F) in a subdivided animal population ($m_c = 0$), as functions of the number of migrants per deme, $N(m_m + m_f)$. (A and D) Equal male and female migration. (B and E) Predominantly male migration ($m_m = 6m_f$). (C and F) Predominantly female migration ($m_f = 6m_m$). We assume $N_{cA} = N$, $N_{cX} = \frac{3}{4}N$, $N_{cY} = N_{cc} = \frac{1}{4}N$.

predominantly female migration where F_{STY} increases much faster with reduced migration than F_{STA} or F_{STX} (Figure 2C). Consequently R_{AY} and R_{XY} show a greater reduction when migration is restricted (Figure 2, D and F) than in the plant model (Figure 1, D and F). These differences are easily understood since *Y*-linked genes now experience migration through only 1 unit (male individuals), while they had two in the plant model (pollen and seeds).

The effects of sex- or chromosome-specific mutation rates: Under the infinite-sites model, the mutation rate is so low compared with any realistic migration rate that differences in mutation rates among different modes of inheritance will not affect the relative F_{ST} values. Absolute values of diversities will, however, be affected, since the ratios for different modes of inheritance will be multiplied by the ratios of the respective mutation rates. From Equations 23, this does not affect the shapes of the R values as functions of migration rates, but does affect their heights. For example, if the male mutation rate is higher than the female mutation rate (Hurst and Ellegren 1998; McVean 2000), this will have the effect of reducing R_{XA}, R_{XY}, and R_{AY}. This would enhance the effect of high relative female migration rates in the animal model on these ratios, which has just been noted.

The effects of nonrandom variation in fertility: It is well known that an increase in the variance of fertility reduces N_e. Sex differences in the distributions of fertility modify the relative N_e values of genes with different modes of inheritance (Caballero 1995; Charlesworth 2001). For example, an increased variance in male fertility reduces the effective deme size of the *Y* chromosome more drastically than that for the *X* or the autosomes. With an extremely high variance in male fertility relative to female fertility, the ratios N_{eX}/N_{eY} and N_{eA}/N_{eY} become 9 and 8, respectively, so that N_{eX}/N_{eA} is 9/8. In contrast, a high variance in female fertility reduces N_e for the *X* and the autosomes but does not affect the *Y* chromosome: the ratios N_{eX}/N_{eY} and N_{eA}/N_{eY} tend to zero and N_{eX}/N_{eA} to 9/16.

As far as total diversity measures are concerned, population subdivision always counteracts the effect of an increased variance in male fertility by reducing the ratios R_{XY} and R_{AY}, as illustrated in the animal model with large ΔV_m (Figure 3). Since N_{eY} is reduced by an increased male fertility variance, the effective number of *Y* migrants is much lower than the number of *X* or autosomal migrants. As the migration rate decreases, $F_{ST,Y}$ now increases much faster than $F_{ST,X}$ or $F_{ST,A}$, and the ratios R_{XY} and R_{AY} decrease toward one with equal male and female migration (Figure 3, B and E). With predominantly female migration, this effect is manifest even with relatively high migration (Figure 3F). For instance, R_{XY} = 6.2 and R_{AY} = 5.7 with very high migration, but are already halved with as many as 10 effective migrants. With an increased variance in female fertility, the effect of population subdivision depends on both the relative effective population sizes and the relative migration rates for different modes of inheritance. Population subdivision increases R_{XY} (R_{AY}) if $N_{eX}\cdot m_X < N_{eY}\cdot m_Y$ ($N_{eA}\cdot m_A < N_{eY}\cdot m_Y$). However, this effect occurs only with very restricted migration, as illustrated in the plant model with $\Delta V_f = 5$ (Figure 4). Moreover, the maximum value of R_{XY} (R_{AY}) is always lower than the expectation for a panmictic population with Poisson distributions of offspring numbers.

Selection on a nonrecombining genome, such as organelle genomes and the *Y* chromosome, can further reduce N_e (Charlesworth and Charlesworth 2000). Such a reduction of N_{ec} and N_{ey} will increase the corresponding F_{ST} values. We have investigated the case with $N_{eX}/N_{eY} = 30$ for the plant model (Figure 5), since this is the ratio suggested by observed within-deme diversity ratios in *Silene latifolia* (see below). With such a low N_{eY}, the decrease in R_{XY} and R_{AY} due to population subdivision now occurs even with relatively high migration rates. Moreover, with such a low N_{eY}, differences between pollen and seed dispersal rates have little effect on the relative total diversities of nuclear genes with different modes of inheritance.

COMPARISONS WITH DATA ON NATURAL POPULATIONS

In this section, we compare some of the results derived above with data from surveys of DNA sequence variation in populations of humans and plants.

Human populations: There is a large literature on genetic diversity in humans, and many different types of markers have been employed (including protein polymorphisms, restriction fragment length polymorphisms, microsatellites, *Alu* insertions, and single-nucleotide polymorphisms). These data have, however, several biases, which limit their utility for our purposes. First, worldwide population structure has rarely been investigated using markers with different modes of inheritance in the same samples (but see Poloni *et al.* 1997; Jorde *et al.* 2000). In addition, the mode of sampling from different populations, or of grouping the populations, may influence F_{ST} estimates, as described by Stoneking (1998) and Hammer *et al.* (2001; see Table 3). Moreover, different estimators of F_{ST}, which depend in different ways on the sample sizes and numbers of populations sampled (Cockerham and Weir 1993; Charlesworth 1998), have been used in different studies, making it hard to compare different results. This is illustrated by the different estimates of $F_{ST,Y}$ obtained from the same dataset by Seielstad *et al.* (1998) and by Hammer *et al.* (2001). Finally, because of the low level of sequence diversity in humans, investigators have favored the use of highly polymorphic markers, especially for the *Y* chromosome. This may bias F_{ST} estimates, as noted by Nagylaki (1998b) and Bertranpetit (2000), since diversity

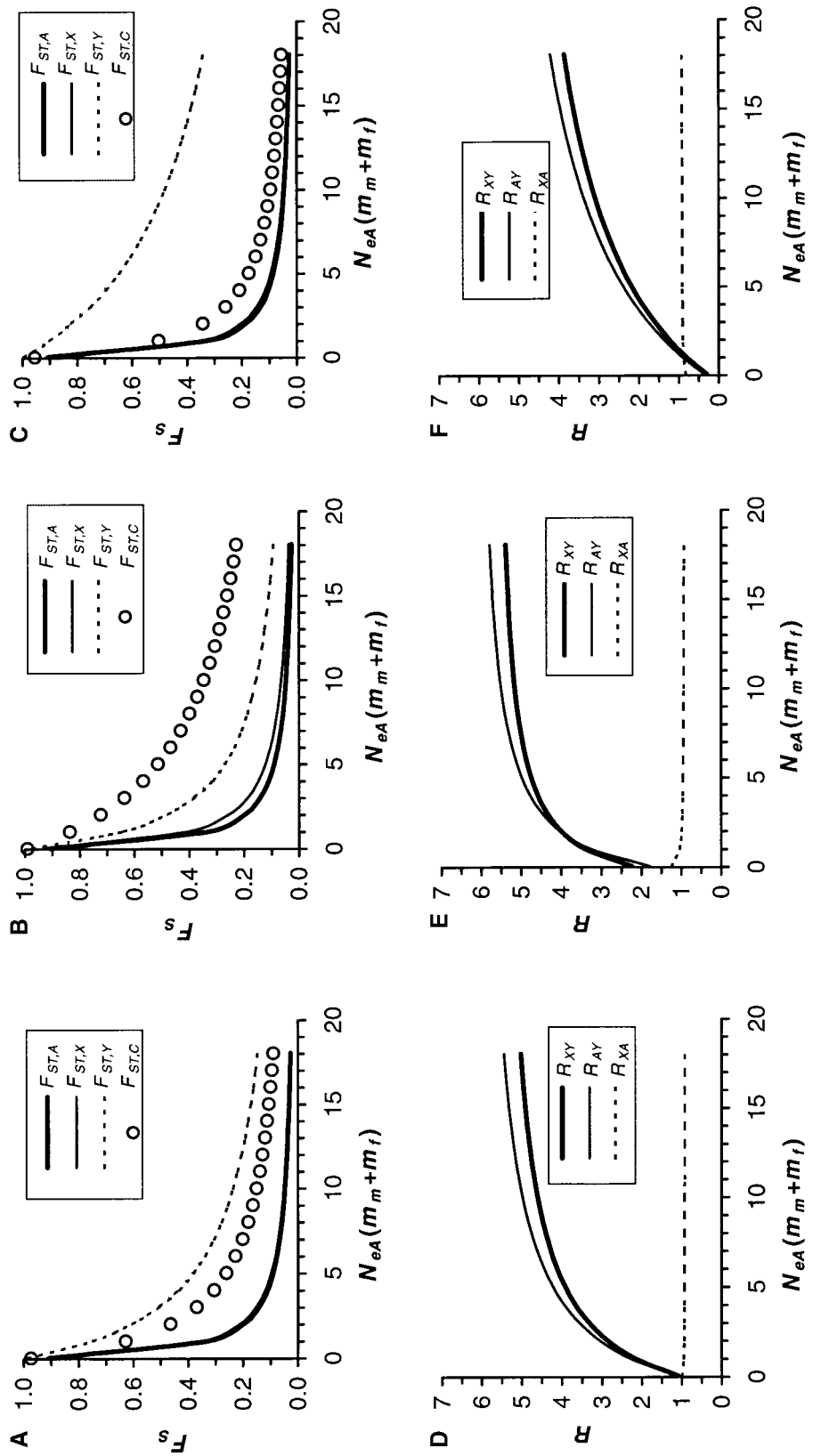

FIGURE 3.—Same as Figure 2 but assuming an increased variance of male fertility compared to the Poisson expectation ($\Delta V_m = 10$).

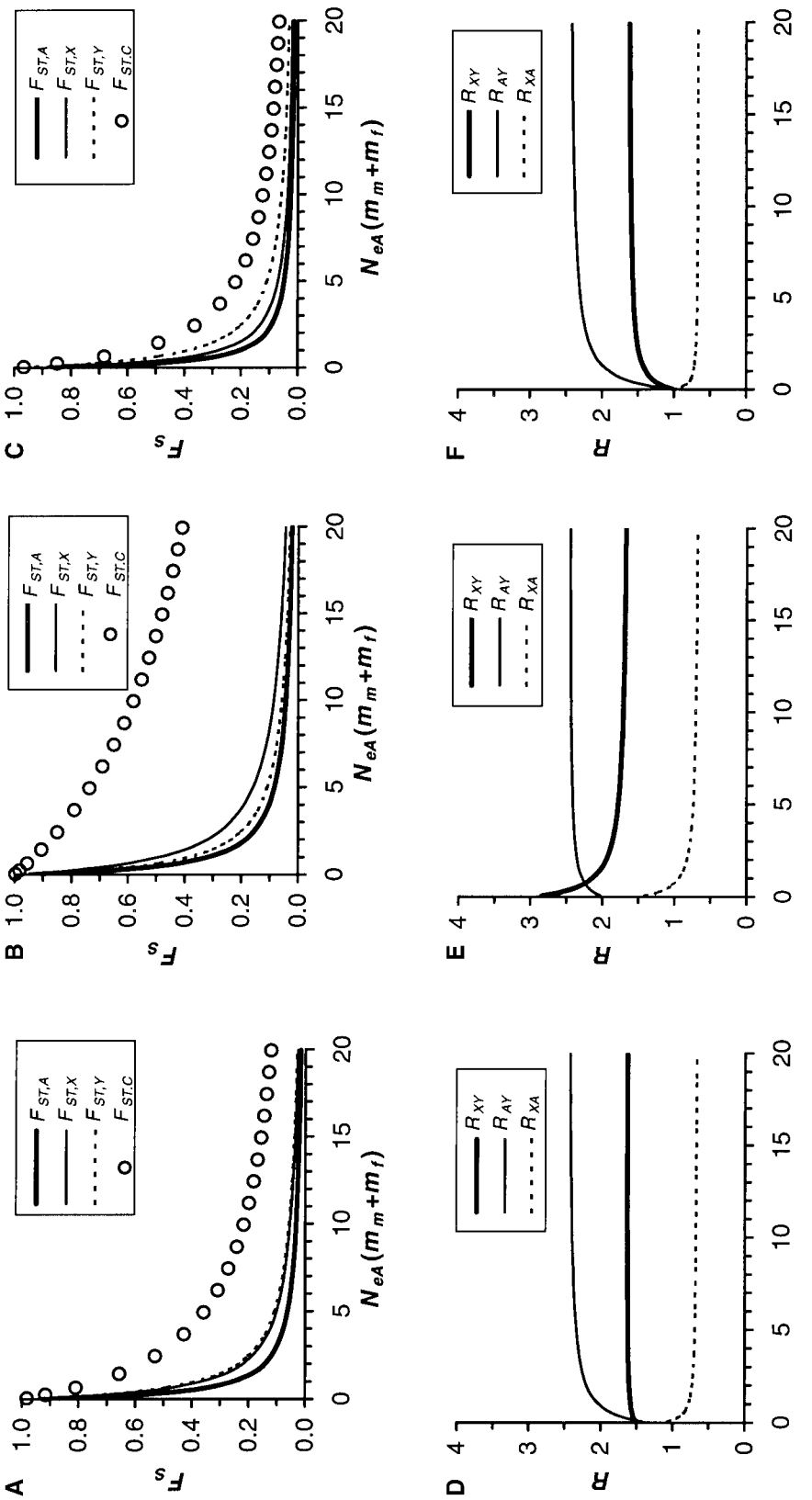

FIGURE 4.—Same as Figure 1 but assuming an increased variance of female fertility compared to the Poisson expectation ($\Delta V_f = 5$).

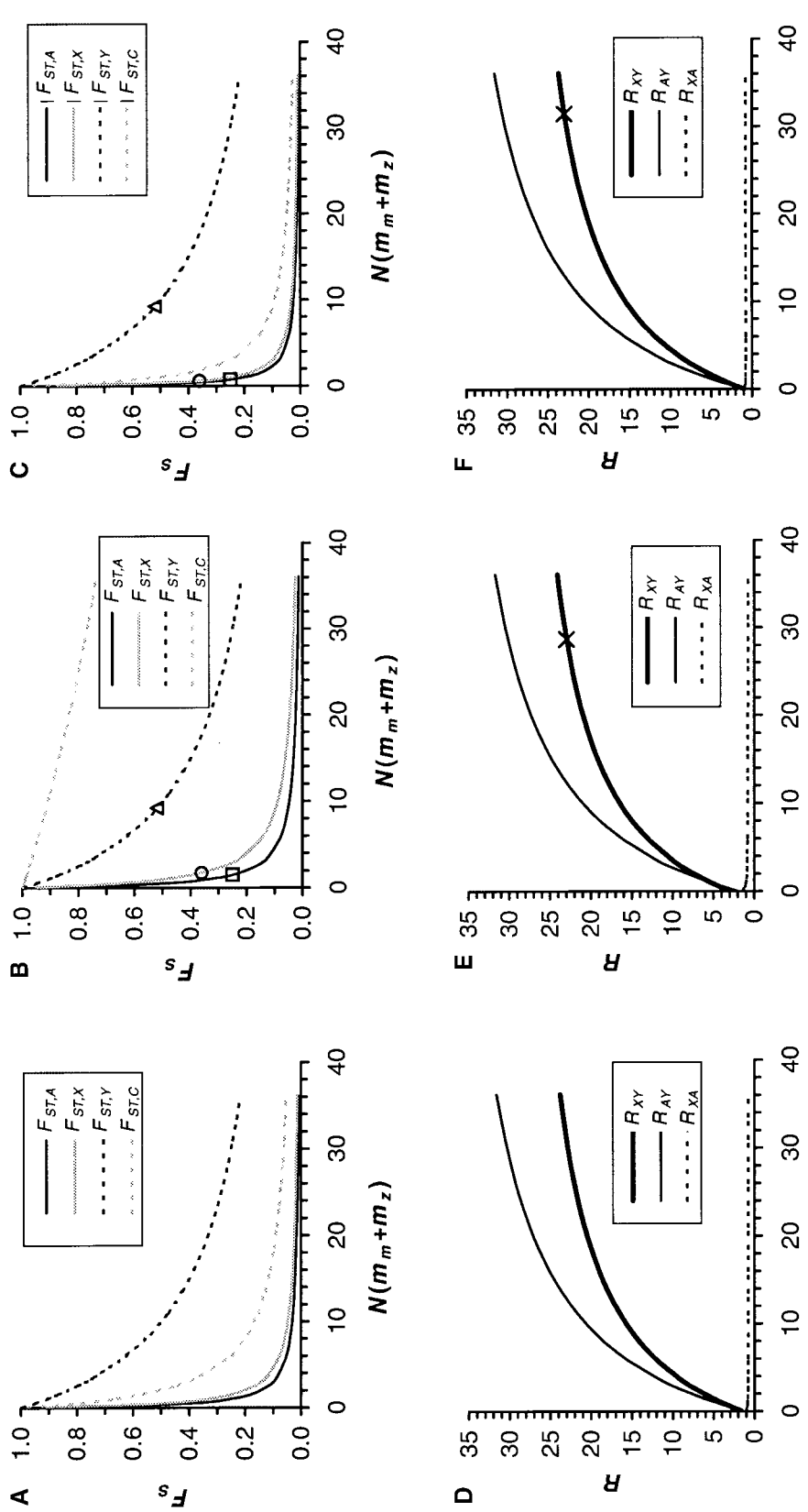

FIGURE 5.—Same as Figure 1 but assuming $N_{eY} = N_{eX}/30$. The figures also give the F_{ST} and R_{XY} estimates from *Silene latifolia* populations with the number of migrants expected under an island model (Triangle, $\hat{F}_{ST,Y}$; square, $\hat{F}_{ST,A}$; circle, $\hat{F}_{ST,X}$; and cross, $\hat{R}_{XY}$).

TABLE 3

Review of F_{ST} estimates at a worldwide spatial scale in human populations

No. of groups or populations	F_{ST} statistics	Gene heredity			Reference
		Y	A	MtDNA	
15	G_{ST} (1)	0.18[a]; 0.54[b]	0.10[a]; 0.11[b]; 0.15[c]	0.25[d]	Jorde *et al.* (2000)
	ψ_{ST} (2)	0.17[a]	0.12[a]; 0.15[b]; 0.19[c]	0.30[d]	
19	ψ_{ST} (2)	0.19[e]	—	0.27[f]	Poloni *et al.* (1997)
15	ψ_{ST} (2)	0.16[a]; 0.30[e]	—	—	Hammer *et al.* (1997)
50	ψ_{ST} (2)	0.36[g]	—	—	Hammer *et al.* (2001)
10	F_{ST} (1)	0.41[h]	—	—	Reviewed in Hammer *et al.* (2001), data from Underhill *et al.* (1997)
	ψ_{ST} (2)	0.54[h]			
—	ψ_{ST} (2)	0.65[h]	—	—	Seielstad *et al.* (1998) from Underhill *et al.* (1997)
20	ψ_{ST} (2)	0.23[a]	—	—	Kayser *et al.* (2001)
21	ψ_{ST} (2)	0.09[a]; 0.40[c]	—	—	Quintana-Murci *et al.* (1999)
5	K_{ST} (3)	0.14[d]	—	—	Thomson *et al.* (2000)
10; 14	ψ_{ST} (2)	—	0.16[e]; 0.16[a]	—	Barbujani *et al.* (1997)
5	F_{ST} (1)	—	0.14[e]	—	Bowcock *et al.* (1991)
31	F_{ST} (1)	—	0.112[c]	—	Watkins *et al.* (2001)
3	Equation 24	—	0.034[d,i]	—	Yu *et al.* (2001)
4	Equation 24	—	0.165[d,i]	—	Zhao *et al.* (2000)
13	F_{ST} (1)	—	0.147[j,k]	—	Zietkiewicz *et al.* (1998)
10	ψ_{ST} (2)	—	—	0.25[f]	Excoffier *et al.* (1992)

(1) F_{ST} or G_{ST}, Cockerham and Weir (1993); Wright (1951); (2) ψ_{ST}, Excoffier *et al.* (1992); (3) K_{ST} Hudson *et al.* (1992).

[a] Microsatellites.
[b] Restriction site polymorphism (RSP).
[c] *Alu*.
[d] Sequences.
[e] RFLP.
[f] mtDNA RFLP.
[g] SSCP + DHPLC.
[h] Denaturing high-performance liquid chromatography (DHPLC).
[i] Introns or noncoding regions.
[j] SSCP + heteroduplex analysis + sequencing.
[k] Neutral polymorphisms.

values for loci with high mutation rates are not proportional to coalescent times.

In Table 3, we present a compilation of F_{ST} estimates from worldwide samples of different types of markers, with cytoplasmic, *Y*, *X*, and autosomal inheritance. These all provide evidence for population structure, with autosomal markers yielding the lowest mean F_{ST} estimates, as expected from the results shown in Figure 2. The estimates obtained from *Y* chromosome markers are, however, highly variable between studies (0.09–0.65). Microsatellites often show smaller F_{ST} values (0.09–0.23), as outlined in studies that have compared both kinds of markers (Hammer *et al.* 1997; Jorde *et al.* 2000), probably because of their high mutation rates. Excluding microsatellites, the average F_{ST} estimated from the *Y* chromosome markers (0.40) is higher than from the mitochondrial markers (0.27). While this has often been interpreted as evidence for higher female than male migration rates, it might also reflect differences in effective population sizes (Hurles and Jobling 2001). A recent comparison of local population differentiation in Thai communities with patrilocal or matrilocal dispersal patterns is, however, in agreement with the expectation that restricted male migration increases population differentiation for *Y* markers relative to mitochondrial markers (Oota *et al.* 2001).

DNA sequence data have recently been obtained for 4 *Y*-linked and 15 autosomal loci, using similar worldwide samples (Shen *et al.* 2000; Thomson *et al.* 2000; Thorstenson *et al.* 2001). The authors estimated R_{AY} as five, from data on both coding and noncoding regions. Unfortunately, population structure was not investigated explicitly in these studies, although the data should allow partitioning of diversity between and within populations. Thomson *et al.* (2000) constructed a genealogy of their sample of *Y* chromosomes, and this can be used to obtain a value of 0.14 for $F_{ST,Y}$ between continents. This is smaller than that obtained from other studies

TABLE 4

Silent site diversity in *Silene latifolia*

Mode of inheritance	Gene (no. of silent sites)	S	$\hat{\pi}_S$ (%)	$\hat{\pi}_T$ (%)	$\hat{F}_{ST}$
Y-linkage	*SlY1* (1652)	4	0.048	0.088	0.45
	SlY4 (829)	5	0.073	0.189	0.62
	SlY1 + *SlY4* (2481)	9	0.056	0.113	0.50
X-linkage	*SlX1* (1262)	88	1.567	2.172	0.28
	SlX4 (362)	42	1.933	3.974	0.51
	SlX1 + *SlX4* (1624)	78	1.649	2.574	0.36
Autosomal	*CCLS37.1* (955)	22	0.610	0.743	0.18

Diversity was estimated from a sample of 13 individuals from five populations: one Danish ($n = 3$), one U.S. (Van, $n = 3$), one Portuguese (SC, $n = 3$), and two British (Dal, $n = 2$; Art, $n = 2$) except for the autosomal gene where the Danish population was not studied (Filatov *et al.* 2001; V. Laporte, unpublished data). F_{ST} was estimated following Equation 24.

of biallelic markers and challenges the notion that there is restricted male *vs.* female migration in most human populations. Mitochondrial data for the same samples would help to resolve this question, and it would also be desirable to obtain estimates of $F_{ST,A}$ and r_{AY}.

At present, it is hard to reach firm conclusions about the influence of sex-specific migration *vs.* differences in effective population sizes on the relative levels of diversity and divergence between human populations for different inheritance modes; there does not, however, seem to be strong evidence for a greatly reduced effective size of the human *Y* chromosome, in contrast to what is observed in Silene (see below) or in Drosophila (Zurovcova and Eanes 1999; Bachtrog and Charlesworth 2000, 2002).

Plant populations: There are relatively few species of plants with sex chromosomes, and diversity data on nuclear genes with different modes of inheritance are available only for the close relatives *S. latifolia* and *S. dioica* (Table 4). In *S. dioica*, there is only a single polymorphic site on the *Y*, which is too little to provide any useful information. In *S. latifolia*, nine polymorphic sites were found on the *Y*. All nuclear genes display quite strong population structure, with the lowest F_{ST} for the autosomal gene and the highest for a *Y*-linked gene, as expected from the above analyses. The ratio R_{XY} over all demes ($R_{XY} = 23$) is smaller than the ratio r_{XY} for within-deme diversity ($r_{XY} = 29$), as expected from the effects of subdivision (Figure 5).

It is, however, impossible to reconcile the estimates of F_{ST} for all three modes of heredity under a simple island model of population subdivision: given the estimates of $F_{ST,A}$ and $F_{ST,X}$, $F_{ST,Y}$ is expected to be much higher (and R_{XY} to be much lower) than is observed (Figure 5). Deviations from the model assumed here (*e.g.*, the occurrence of a selective sweep on the *Y*) might account for these discrepancies, if they are not simply due to sampling error due to the small number of informative sites on the *Y*.

DISCUSSION

In the first part of this article, we have shown that the use of the fast-timescale approximation (Nordborg 1997) to the coalescent process in a single population, structured according to sex and/or age, provides a simple general expression for the probability that a pair of alleles sampled from the population coalesce in the previous time interval (Equation 6). This approximation assumes that the flow of genes among age and sex classes is much faster than the effects of drift or migration between populations, so that the state of an allele can be treated as independent of the class from which it was sampled: this is equivalent to the "strong-migration limit" of Nagylaki (1980) for a geographically structured population.

By use of a suitable definition of generation time, we can also define the effective size of a population as the reciprocal of twice the product of generation time and the probability of coalescence per time interval (Equation 7). This expression can be used to generate explicit formulas for effective population size with discrete generations and separate sexes, under a variety of different modes of inheritance, and with arbitrary distributions of offspring numbers (Equations 14–17). The same approach can be used for the case of a nuclear gene in a population of partially self-fertilizing hermaphrodites (Equations 19 and 20). An important conclusion in this case is that the standard formula for N_e for selfing populations (Pollak 1987) is likely to be approximately valid even if there is a non-Poisson distribution of offspring number; *i.e.*, N_e is reduced below the value for random mating by a factor of $1/(1 + F)$, where F is the inbreeding coefficient.

The approach can also be applied to the standard model of an age-structured population with discrete time intervals, recently revisited by Charlesworth (2001); his Equation A7 is equivalent to our Equations 6 and 7. Finally, we note that the effective population size defined in this way is equivalent to that which describes

the asymptotic rate of increase in the probability of identity of a pair of randomly sampled genes in the absence of mutation, *i.e.*, to the inbreeding effective population size. This equivalence is dependent on the validity of the fast-timescale approximation (Nagylaki 1980; Charlesworth 2001).

We also show how to simplify the analysis of the effects of population subdivision of a demographically structured population by defining a migration matrix that describes the net rates of movement of genes between different local populations. This involves weighting the migration probabilities of individuals of a given age-sex class by the contribution of this class to the leading left eigenvector of the matrix describing movements of genes between age and sex classes (Equation 4). This enables the determination of the moment-generating functions for the distributions of coalescent times for pairs of genes sampled from a given pair of populations, under any well-defined migration model (Equation 5), under the standard assumption that migration and drift are both weak evolutionary forces. From these, the expected coalescent times and higher moments can easily be found (Hudson 1990; Wilkinson-Herbots 1998).

Under the infinite-sites model, the expected number of nucleotide differences between a pair of alleles sampled from a prescribed pair of populations is equal to the product of the mutation rate and the corresponding expected coalescent time (Hudson 1990), so that the results can be used to generate expressions for equilibrium levels of nucleotide site diversities. In doing this, the mutation rates for different age-sex classes must be weighted in the same way as the migration rates. This approach can be extended to extinction-recolonization metapopulation models, if it is assumed that the extinction of a local population is immediately followed by its recolonization (Wakeley and Aliacar 2001).

An important general conclusion is that population subdvision makes it very hard to describe the expected level of genetic variability in a population by a single formula such as $\theta = 4N_e u$. While it is possible to define a simple expression for a weighted mean coalescence time for a pair of alleles sampled from the same deme for a general migration model (Equation 9), this involves both the effective population sizes of all the individual demes and their contributions to the leading left eigenvector of the migration matrix defined by Equation 4. In general, these are unknowable quantities, making it very hard to equate any empirical estimate of the mean within-population nucleotide site diversity to a simple scaled mutation rate parameter. The details of the demography and migration parameters of a species may greatly influence the estimated scaled mutation rate based on unweighted mean within-population nucleotide site diversities, making comparisons between different species difficult to interpret.

The situation is even worse for the nucleotide diversity for a pair of alleles sampled from the population at large, since this is related to the within-deme value by $1/(1 - F_{ST})$. In general, F_{ST} depends in a complex way on migration rates and deme sizes, and simple formulas are available for only a few limiting cases, such as the island model. Only when there is negligible genetic differentiation between local populations can one confidently relate mean nucleotide site diversity to the coalescent time for a randomly sampled pair of alleles and hence to a scaled mutation rate parameter. Many empirical investigations of DNA sequence variation in natural populations do not state explicitly what population parameters are of primary interest and often make little distinction between measures of variation based on whole-population and within-population estimates. More attention to these issues in presenting analyses of data on DNA sequence variability is desirable.

Our investigation of the island model also shows that strong population subdivision, coupled with sex-specific migration rates, can greatly affect the relative values of the expected total genetic diversities for different modes of inheritance and may even reverse some of the patterns expected under panmixia. For example, Figure 3 displays the expected patterns of genetic variability for animal populations with a high variance in male reproductive success. From Figure 3C, it may be seen that predominantly female migration can cause the ratios of autosomal or *X*-linked variability to *Y*-linked variability to decline below one with extreme subdivision, in contrast to a value of over four with panmixia. Encouragingly, however, the relative values of expected autosomal and *X*-linked total population diversities are insensitive to all but the most extreme population subdivision, for both plant and animal models (Figures 1–4). This suggests that the use of ratios of *X*-linked to autosomal diversities to make inferences about the strength of sexual selection (Charlesworth 2001) may be feasible even if there is a high level of population subdivision.

We have confined ourselves to deriving expressions for expected coalescence times and nucleotide site differences between alleles. However, expressions for the distribution of coalescent times for a set of n alleles sampled from a specified set of populations can easily be written down, using the migration probabilities defined by Equation 4 and coalescence probabilities defined by Equation 7 to generate the expectations of competing exponential distributions of waiting times to migration or coalescence events (Hudson 1990). Although in general it is difficult to obtain useful analytical results on the properties of these distributions (but see Wakeley 1999, 2001), they can readily be implemented in computer simulations and used to derive sampling properties of estimators of the type that we have discussed.

We thank N. Barton, D. Charlesworth, F. Depaulis, and two anonymous reviewers for their comments on the manuscript. B.C. acknowledges support by the Royal Society and the Engineering and Physics

Research Council, and V.L. acknowledges support by the Biotechnology and Biological Sciences Research Council.

LITERATURE CITED

Bachtrog, D., and B. Charlesworth, 2000 Reduced levels of microsatellite variability on the neo-*Y* chromosome of Drosophila miranda. Curr. Biol. **10:** 1025–1031.

Bachtrog, D., and B. Charlesworth, 2002 Reduced adaptation of an evolving neo-*Y* chromosome. Nature **416:** 323–326.

Barbujani, G., A. Magagni, E. Minch and L. L. Cavalli-Sforza, 1997 An apportionment of human DNA diversity. Proc. Natl. Acad. Sci. USA **94:** 4516–4519.

Bertranpetit, J., 2000 Genome, diversity, and origins: the *Y* chromosome as a storyteller. Proc. Natl. Acad. Sci. USA **97:** 6927–6929.

Bowcock, A. M., J. R. Kidd, J. L. Mountain, J. M. Hebert, L. Carotenuto *et al.*, 1991 Drift, admixture, and selection in human evolution—a study with DNA polymorphisms. Proc. Natl. Acad. Sci. USA **88:** 839–843.

Caballero, A., 1995 On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics **139:** 1007–1011.

Charlesworth, B., 1998 Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. **15:** 538–543.

Charlesworth, B., 2001 The effect of life-history and mode of inheritance on neutral genetic variability. Genet. Res. **77:** 153–166.

Charlesworth, B., and D. Charlesworth, 2000 The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. B **355:** 1563–1572.

Charlesworth, D., and B. Charlesworth, 1981 Allocation of resources to male and female functions in hermaphrodites. Biol. J. Linn. Soc. **15:** 57–74.

Chesser, R. K., and R. J. Baker, 1996 Effective sizes and dynamics of uniparentally and diparentally inherited genes. Genetics **144:** 1225–1235.

Chesser, R. K., O. E. J. Rhodes, D. W. Sugg and A. Schnabel, 1993 Effective sizes for subdivided populations. Genetics **135:** 1221–1232.

Cockerham, C. C., and B. S. Weir, 1993 Estimation of gene flow from *F*-statistics. Evolution **47:** 855–863.

Ennos, R. A., 1994 Estimating the relative rates of pollen and seed migration among plant populations. Heredity **72:** 250–259.

Excoffier, L., P. E. Smouse and J. M. Quattro, 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes—application to human mitochondrial-DNA restriction data. Genetics **131:** 479–491.

Filatov, D. A., F. Monéger, I. Negrutiu and D. Charlesworth, 2000 Evolution of a plant *Y*-chromosome: variability in a *Y*-linked gene of *Silene latifolia*. Nature **404:** 388–390.

Filatov, D. A., V. Laporte, C. Vitte and D. Charlesworth, 2001 DNA diversity in sex linked and autosomal genes of the plant species *Silene latifolia* and *S. dioica*. Mol. Biol. Evol. **18:** 1442–1454.

Hammer, M. F., A. B. Spurdle, T. Karafet, M. R. Bonner, E. T. Wood *et al.*, 1997 The geographic distribution of human *Y* chromosome variation. Genetics **145:** 787–805.

Hammer, M. F., T. M. Karafet, A. J. Redd, H. Jarjanazi, S. Santachiara-Benerecetti *et al.*, 2001 Hierarchical patterns of global human *Y*-chromosome diversity. Mol. Biol. Evol. **18:** 1189–1203.

Hudson, R. R., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7:** 1–45.

Hudson, R. R., D. D. Boos and N. L. Kaplan, 1992 A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:** 138–151.

Hurles, M. E., and M. A. Jobling, 2001 Haploid chromosomes in molecular ecology: lessons from the human *Y*. Mol. Ecol. **10:** 1599–1613.

Hurst, L. D., and H. Ellegren, 1998 Sex biases in the mutation rate. Trends Genet. **14:** 446–452.

Jorde, L. B., W. S. Watkins, M. J. Bamshad, M. E. Dixon, C. E. Ricker *et al.*, 2000 The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and *Y*-chromosome data. Am. J. Hum. Genet. **66:** 979–988.

Kayser, M., M. Krawczak, L. Excoffier, P. Dieltjes, D. Corach *et al.*, 2001 An extensive analysis of *Y*-chromosomal microsatellite haplotypes in globally dispersed human populations. Am. J. Hum. Genet. **68:** 990–1018.

Kimura, M., 1971 Theoretical foundations of population genetics at the molecular level. Theor. Popul. Biol. **2:** 174–208.

Lloyd, D. G., 1977 Genetic and phenotypic models of natural selection. J. Theor. Biol. **69:** 543–560.

McVean, G., 2000 Evolutionary genetics: what is driving male mutation? Curr. Biol. **10:** R834–R835.

McVean, G. A. T., and L. D. Hurst, 1997 Evidence for a selectively favourable reduction in the mutation rate of the *X* chromosome. Nature **386:** 388–392.

Nagylaki, T., 1980 The strong migration limit in geographically structured populations. J. Math. Biol. **9:** 101–114.

Nagylaki, T., 1982 Geographical invariance in population genetics. J. Theor. Biol. **99:** 159–172.

Nagylaki, T., 1998a The expected number of heterozygous sites in a subdivided population. Genetics **149:** 1599–1604.

Nagylaki, T., 1998b Fixation indices in subdivided populations. Genetics **148:** 1325–1332.

Nordborg, M., 1997 Structured coalescent processes on different time scales. Genetics **146:** 1501–1514.

Nordborg, M., and P. Donnelly, 1997 The coalescent process with selfing. Genetics **146:** 1185–1195.

Oota, H., W. Settheetham-Ishida, D. Tiwawech, T. Ishida and M. Stoneking, 2001 Human mtDNA and *Y*-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat. Genet. **29:** 20–21.

Pollak, E., 1987 On the theory of partially inbreeding populations. I. Partial selfing. Genetics **117:** 353–360.

Poloni, E. S., O. Semino, G. Passarino, A. S. Santachiara-Benerecetti, I. Dupanloup *et al.*, 1997 Human genetic affinities for *Y*-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. Am. J. Hum. Genet. **61:** 1015–1035.

Quintana-Murci, L., O. Semino, E. S. Poloni, A. Liu, M. Van Gijn *et al.*, 1999 *Y*-chromosome specific YCAII, DYS19 and YAP polymorphisms in human populations: a comparative study. Ann. Hum. Genet. **63:** 153–166.

Rousset, F., 1999 Genetic differentiation in populations with different classes of individuals. Theor. Popul. Biol. **55:** 297–308.

Seielstad, M. T., E. Minch and L. L. Cavalli-Sforza, 1998 Genetic evidence for a higher female migration rate in humans. Nat. Genet. **20:** 278–280.

Shen, P. D., F. Wang, P. A. Underhill, C. Franco, W. H. Yang *et al.*, 2000 Population genetic implications from sequence variation in four *Y* chromosome genes. Proc. Natl. Acad. Sci. USA **97:** 7354–7359.

Slatkin, M., 1991 Inbreeding coefficients and coalescence times. Genet. Res. **58:** 167–175.

Stoneking, M., 1998 Women on the move. Nat. Genet. **20:** 219–220.

Thomson, R., J. K. Pritchard, P. D. Shen, P. J. Oefner and M. W. Feldman, 2000 Recent common ancestry of human *Y* chromosomes: evidence from DNA sequence data. Proc. Natl. Acad. Sci. USA **97:** 7360–7365.

Thorstenson, Y. R., P. D. Shen, V. G. Tusher, T. L. Wayne, R. W. Davis *et al.*, 2001 Global analysis of ATM polymorphism reveals significant functional constraint. Am. J. Hum. Genet. **69:** 396–412.

Underhill, P. A., L. Jin, A. A. Lin, S. Q. Mehdi, T. Jenkins *et al.*, 1997 Detection of numerous *Y* chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res. **7:** 996–1005.

Wakeley, J., 1999 Nonequilibrium migration in human history. Genetics **153:** 1863–1871.

Wakeley, J., 2001 The coalescent in an island model of population subdivision with variation among demes. Theor. Popul. Biol. **59:** 133–144.

Wakeley, J., and N. Aliacar, 2001 Gene genealogies in a metapopulation. Genetics **159:** 893–905.

Wang, J., 1997a Effective size and *F*-statistics of subdivided populations: I. Monoecious species with partial selfing. Genetics **146:** 1453–1463.

Wang, J., 1997b Effective size and *F*-statistics of subdivided populations: II. Dioecious species. Genetics **146:** 1465–1474.

Wang, J., 1999 Effective size and *F*-statistics of subdivided populations for sex-linked loci. Theor. Popul. Biol. **55:** 176–188.

Wang, J., and A. Caballero, 1999 Developments in predicting the effective size of subdivided populations. Heredity **82:** 212–226.
Watkins, W. S., C. E. Ricker, M. J. Bamshad, M. L. Carroll, S. V. Nguyen *et al.*, 2001 Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. Am. J. Hum. Genet. **68:** 738–752.
Whitlock, M. C., and N. H. Barton, 1997 The effective size of a subdivided population. Genetics **146:** 427–441.
Wilkinson-Herbots, H. M., 1998 Genealogy and subpopulation differentiation under various models of population structure. J. Math. Biol. **37:** 535–585.
Wright, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97–169.
Wright, S., 1943 Isolation by distance. Genetics **28:** 114–138.
Wright, S., 1951 The genetical structure of populations. Ann. Eugen. **15:** 323–354.
Yu, N., Y. X. Fu, N. Sambuughin, M. Ramsay, T. Jenkins *et al.*, 2001 Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. Mol. Biol. Evol. **18:** 214–222.
Zhao, Z. M., L. Jin, Y. X. Fu, M. Ramsay, T. Jenkins *et al.*, 2000 Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc. Natl. Acad. Sci. USA **97:** 11354–11358.
Zietkiewicz, E., V. Yotova, M. Jarnik, M. Korah-Laskowska, K. K. Kidd *et al.*, 1998 Genetic structure of the ancestral population of modern humans. J. Mol. Evol. **47:** 146–155.
Zurovcova, M., and W. F. Eanes, 1999 Lack of nucleotide polymorphism in the *Y*-linked sperm flagellar dynein gene *Dhc-Yh3* of *Drosophila melanogaster* and *D. simulans*. Genetics **153:** 1709–1715.

Communicating editor: G. B. Golding

APPENDIX

Variances in offspring numbers with a binomial distribution of the proportion of sons *vs.* daughters: Let the proportion of males among breeding individuals be c. From the properties of the binomial distribution, it is easily seen that the variance of the number of sons per female is

$$V_{\mathrm{fm}} = c^2 V_{\mathrm{f}} + \bar{n}_{\mathrm{f}} c(1 - c), \tag{A1a}$$

where $\bar{n}_{\mathrm{f}}$ and V_{f} are the mean and variance of total offspring number per female. If the population size is stationary, we have $\bar{n}_{\mathrm{f}} = 1/(1 - c)$, and so

$$V_{\mathrm{fm}} = c^2 V_{\mathrm{f}} + c. \tag{A1b}$$

The Poisson expectation for V_{fm} is $c/(1 - c)$, so this yields

$$\Delta V_{\mathrm{fm}} = \frac{c^2\{(1 - c)V_{\mathrm{f}} - 1\}}{(1 - c)}. \tag{A1c}$$

Similarly, we have

$$\Delta V_{\mathrm{ff}} = (1 - c)\{(1 - c)V_{\mathrm{f}} - 1\} \tag{A2}$$

$$\Delta C_{\mathrm{ffm}} = c(1 - c)V_{\mathrm{f}} - c. \tag{A3}$$

Corresponding expressions can be derived for the progeny of male parents, interchanging c and $(1 - c)$.

Effective population size with self-fertilization: Consider first two maternally derived genes, sampled from different individuals (probability one-quarter). Their probability of origin from a common parent in the previous generation is independent of whether the individuals are the products of selfing or outcrossing and is simply

$$Q_{\mathrm{ff}} = \frac{(\Delta V_{\mathrm{f}}/\bar{d}^2) + 1}{N}, \tag{A4a}$$

where $\bar{d} = \bar{d}_{\mathrm{S}} + \bar{d}_{\mathrm{O}}$ is the mean per capita number of successful offspring produced through seed, and ΔV_{f} is the excess over $\bar{d}$ of the variance in the number of successful offspring produced through seed.

Next, consider a maternally derived and a paternally derived gene from different individuals (probability one-half). The latter has a probability of S of being derived by selfing, in which case the probability that it came from the same parent as the maternally derived gene is

$$Q_{\mathrm{fmS}} = \frac{\sum_k d_{k\mathrm{S}}(d_{k\mathrm{O}} + d_{k\mathrm{S}} - 1)}{N^2\,\bar{d}_{\mathrm{S}}\bar{d}} = \frac{(C_{\mathrm{fOfS}} + \Delta V_{\mathrm{fS}})/(\bar{d}_{\mathrm{S}}\bar{d}) + 1}{N}, \tag{A4b}$$

where C_{fOfS} is the covariance between the number of offspring produced through outcrossed and selfed seeds, respectively, and ΔV_{fS} is the excess over $\bar{d}_{\mathrm{S}}$ of the variance in number of offspring produced through selfing.

If the paternal allele was derived by outcrossing (probability $1 - S$), the probability that the pair came from a common parent is

$$Q_{\mathrm{fmO}} = \frac{(C_{\mathrm{fm}}/\bar{d}\bar{p}) + 1}{N}, \tag{A4c}$$

where C_{fm} is the covariance between the total number of offspring produced by seed and the number of offspring produced through outcrossed pollen, and $\bar{p}$ is the mean number of offspring produced through outcrossed pollen.

Finally, consider a pair of paternally derived genes (probability one-quarter). There is a probability of approximately S^2 that they are both products of selfing, in which case the probability of sharing a common parent is

$$Q_{\mathrm{fSfS}} = \frac{(\Delta V_{\mathrm{fS}}/\bar{d}_{\mathrm{S}}^2) + 1}{N}. \tag{A4d}$$

There is a probability $2S(1 - S)$ that one is derived from selfing and the other from outcrossing, in which case the probability of common parentage is

$$Q_{\mathrm{fOfS}} = \frac{(C_{\mathrm{fSm}}/\bar{d}_{\mathrm{s}}\bar{p}) + 1}{N}, \tag{A4e}$$

where C_{fSm} is the covariance between the number of offspring produced through selfed seed and outcrossed pollen.

There is a probability $(1 - S)^2$ that both are products of outcrossing, in which case their probability of common parentage is

$$Q_{\mathrm{mOmO}} = \frac{(\Delta V_{\mathrm{m}}/\bar{p}^2) + 1}{N}, \tag{A4f}$$

where ΔV_{m} is the excess over $\bar{p}$ of the variance in number of offspring produced through outcrossed pollen.

Conditioned on common parentage, each of these possible origins of gene pairs has a probability of $(1 + F)/2$ of resulting in coalescence; the reciprocal of the effective population size, N_{eH}, is thus given by multiplying $(1 + F)$ by the sum of the products of the probabilities of origins and common parentage conditioned on origin, where at equilibrium under selfing and random mating $F = S/(2 - S)$.

Assume that the kth individual has a total amount of resource R_k available for reproduction, of which a fraction c_k is devoted to pollen production and $1 - c_k$ to seed production. Let the net expected contribution of offspring through seed be a function $f(R_k[1 - c_k])$ of the individual's allocation to seed production. If the population size is stationary, this has a mean of one.

If fitnesses are normalized so that mean fitnesses through male and female contributions are equal (Lloyd 1977), the expected contribution to the next generation through outcrossed pollen is

$$\bar{p}_i = \frac{g(Rc_i)(1 - \mathrm{S})}{\bar{g}}, \tag{A5}$$

where the function g describes the dependence of relative male reproductive success on allocation to pollen (the expectation of g is 1).

The actual numbers of offspring produced are Poisson variates with $\bar{f}$ and $\bar{p}$ as parameters. The excess of the variance over Poisson expectation in the number of offspring produced through seed is

$$\Delta V_{\mathrm{f}} \approx \left(\frac{\partial f}{\partial R}\right)^2 V_R + \left(\frac{\partial f}{\partial c}\right)^2 V_c \tag{A6a}$$

(assuming that R_k and c_k are independent), where the derivatives are evaluated at the population means.

$$\Delta V_{\mathrm{m}} \approx \frac{(1 - S)^2}{\bar{g}^2}\left\{\left(\frac{\partial g}{\partial R}\right)^2 V_R + \left(\frac{\partial g}{\partial c}\right)^2 V_c\right\}. \tag{A6b}$$

Similarly

$$C_{\mathrm{fm}} \approx \frac{(1 - S)}{\bar{g}}\left\{\left(\frac{\partial f}{\partial R}\right)\left(\frac{\partial g}{\partial R}\right)V_R + \left(\frac{\partial f}{\partial c}\right)\left(\frac{\partial g}{\partial c}\right)V_c\right\} \tag{A6c}$$

and

$$C_{\mathrm{fsm}} \approx \frac{S(1 - S)}{\bar{g}}\left\{\left(\frac{\partial f}{\partial R}\right)\left(\frac{\partial g}{\partial R}\right)V_R + \left(\frac{\partial f}{\partial c}\right)\left(\frac{\partial g}{\partial c}\right)V_c\right\}. \tag{A6d}$$

Substituting these equations into Equation 19b, we obtain

$$\frac{1}{N_{\mathrm{eH}}} \approx \frac{(1 + F)}{N}\left\{1 + \frac{1}{4}\left[(1 + S)\frac{\partial f}{\partial R} + (1 - S)\frac{\partial g}{\partial R}\right]^2 V_R + \frac{1}{4}\left[(1 + S)\frac{\partial f}{\partial c} + \frac{(1 - S)}{\bar{g}}\frac{\partial g}{\partial c}\right]^2 V_c\right\}. \tag{A7}$$

The second term in brackets is equal to the square of the derivative of total fitness with respect to sex allocation (Charlesworth and Charlesworth 1981) and is therefore expected to be close to zero in populations that have had time to adjust their sexual allocation to its evolutionary optimum. The effect of non-Poisson variation in fitness on N_{e} is therefore likely to be largely determined by effects on R. It can be seen that the derivative of the female component of fitness with respect to R has a higher weight, the higher the selfing rate, whereas the derivative of outcrossed male fitness has a weight that declines to zero for complete selfing. If the two derivatives are of similar magnitude, there will be very little change in the contribution from V_R with changes in S.

More complex models, which allow for variation in the selfing rate and for covariances between R and c, can be written down, but the conclusions are not greatly changed. The covariance term can be seen to carry a weight that involves a factor equal to the derivative of total fitness with respect to sex allocation, which is expected to be close to zero (see above) from the above argument. The variance in selfing rate is necessarily close to zero for outcrossing and highly selfing populations and so will not influence their relative N_{e} values.